



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Identification of potential SARS-CoV-2 entry inhibitors by targeting the interface region between the spike RBD and human ACE2



Arun Bahadur Gurung^{a,*}, Mohammad Ajmal Ali^b, Joongku Lee^c, Mohammad Abul Farah^d, Khalid Mashay Al-Anazi^d

^a Department of Basic Sciences and Social Sciences, North-Eastern Hill University, Shillong, 793022, Meghalaya, India

^b Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, 11451, Saudi Arabia

^c Department of Environment and Forest Resources, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon, 34134, Republic of Korea

^d Department of Zoology, College of Science, King Saud University, Riyadh, 11451, Saudi Arabia

ARTICLE INFO

Article history:

Received 16 September 2020

Received in revised form 4 November 2020

Accepted 8 December 2020

Keywords:

Receptor-binding domain

RBD

Angiotensin-converting enzyme 2

ACE2

Spike protein

Protein-protein interface

COVID-19

SARS-CoV-2

Molecular docking

Molecular dynamics simulation

ABSTRACT

Coronavirus disease 2019 (COVID-19) is a fatal infectious disease caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). The virus infection is initiated upon recognition and binding of the spike (S) protein receptor-binding domain (RBD) to the host cell surface receptor, angiotensin-converting enzyme 2 (ACE2). Blocking the interaction between S protein and ACE2 receptor is a novel approach to prevent the viral entry into the host cell. The present study is aimed at the identification of small molecules which can disrupt the interaction between SARS-CoV-2 S protein and human ACE2 receptor by binding to the interface region. A chemical library consisting of 1,36,191 molecules were screened for drug-like compounds based on Lipinski's rule of five, Verber's rule and *in silico* toxicity parameters. The filtered drug-like molecules were next subjected to molecular docking in the interface region of RBD. The best three hits viz; ZINC64023823, ZINC33039472 and ZINC00991597 were further taken for molecular dynamics (MD) simulation studies and binding free energy evaluations using Molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) and Molecular mechanics-Generalized Born surface area (MM-GBSA). The protein-ligand complexes showed stable trajectories throughout the simulation time. ZINC33039472 exhibited binding free energy value lower as compared to the control (emodin) with a higher contribution by gas-phase energy and van der Waals energy to the total binding free energy. Thus, ZINC33039472 is identified to be a promising interfacial binding molecule which can inhibit the interaction between the viral S protein and human ACE2 receptor which would consequently help in the management of the disease.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of King Saud Bin Abdulaziz University for Health Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Since its first emergence in December 2019 in Wuhan city, Hubei province of China, the highly pathogenic coronavirus known ini-

Abbreviations: 2019-nCoV, 2019 novel coronavirus; ACE2, Angiotensin-converting enzyme 2; Covid-19, coronavirus disease 2019; DPP4, dipeptidyl peptidase 4; LARMD, ligand and receptor molecular dynamics; MERS-CoV, Middle East respiratory syndrome coronavirus; MM-GBSA, molecular mechanics-generalized born surface area; MM-PBSA, molecular mechanics-Poisson-Boltzmann surface area; PCA, principal component analysis; PDB, Protein Data Bank; PPIs, protein-protein interactions; Q, fraction of native contacts; RBD, receptor-binding domain; Rg, radius of gyration; RMSD, root mean square deviation; ROF, rule of five; SARS-CoV-2, severe acute respiratory syndrome coronavirus-2.

* Corresponding author.

E-mail address: arunbgurung@gmail.com (A.B. Gurung).

<https://doi.org/10.1016/j.jiph.2020.12.014>

1876-0341/© 2020 The Author(s). Published by Elsevier Ltd on behalf of King Saud Bin Abdulaziz University for Health Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

tially as a 2019 novel coronavirus (2019-nCoV) and later described as severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has rapidly spread to more than eighty other countries leading to serious global health emergency [1]. The virus has caused a new fatal disease known as coronavirus disease 2019 (COVID-19) with patients reporting a wide range of symptoms such as fever, dry cough, dyspnea, headache, and pneumonia [2,3]. Two other highly pathogenic human coronaviruses have been previously reported viz; the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012. Compared to the previous outbreaks from these two viruses, SARS-CoV-2 has a low mortality rate which ranges between 3 to 5% [1]. Phylogenetic studies have shown that SARS-CoV-2 is a new member within *betacoronavirus* genus which includes MERS-CoV, SARS-CoV, bat SARS-related coronaviruses (SARSr-CoV) etc.

[4]. The closest relative of the SARS-CoV-2 is most likely to be Bat coronavirus RaTG13 which shows a similarity of 93.1% sequence identity in the spike (S) gene [5]. At present, there are no effective prophylactics or therapeutics available for the treatment of COVID-19 and as of May 5, 2020, there has been a total of 3 489 053 confirmed cases and 241 559 confirmed death reports [6].

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a single positive-strand RNA enveloped virus with a genome size of 29.9 kb [7]. There are four structural proteins present in coronavirus viz; spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins [3]. Out of these, S protein, a type I transmembrane glycoprotein mediates key functions such as viral attachment, fusion and entry [8]. The first step in the viral infection cycle is the viral entry into the host cell which is mediated by binding of the S protein receptor-binding domain (RBD) to the host cell surface receptor which is angiotensin-converting enzyme 2 (ACE2) receptor in SARS-CoV and dipeptidyl peptidase 4 (DPP4) in case of MERS-CoV [9,10]. Similarly, ACE2 also serves as a receptor for SARS-CoV-2 S protein. The RBD (residues 318–510) in the spike S1 subunit (residues 17–680) mediates the binding of the virus to the host cell surface receptor and the transmembrane S2 subunit (residues 681–1195) allows the fusion between the virus and the host cell membrane [11,12]. Because of its significant role in viral entry into the host cell, S protein has emerged as an important therapeutic target for the development of antibodies, vaccines and entry inhibitors [6].

Protein–protein interactions (PPIs) play an important role in regulating various cellular processes as well as drive the pathogenesis in many human diseases such as neurodegenerative diseases, cervical cancer, bacterial infection and leukaemia etc. [13]. Computational study of protein–protein interactions plays a significant role in the drug discovery process which can provide useful insights into the biochemical nature of the protein–protein interface, mutations affecting the binding interface and designing small molecules which can regulate the pathological cellular processes [14]. This rational approach has been widely used in designing of many PPI modulators which can be categorized into three different types— a) orthosteric inhibitor b) allosteric inhibitor and c) interfacial inhibitor [15]. The orthosteric inhibitors bind to the protein target preferentially at regions which are normally used by other interacting protein partners for binding and therefore, disrupt the formation of functional biological complexes. A typical example of orthosteric inhibitor is MI-2-2 which binds to menin with nanomolar binding affinity and disrupts the protein–protein interaction between menin and MLL in mixed lineage leukemia (MLL) leukaemia cells [16]. Whereas allosteric inhibitors bind to sites distinct from the protein–protein interface region and cause a conformational change in the target protein and thereby destabilizing the protein–protein complex. BRAf inhibitor, PLX4032 is an example of an allosteric inhibitor that inhibits BRAf-CRAF heterodimerization and activation [17]. The interfacial inhibitors are small molecules which bind to a pocket at the protein–protein interface to form a ternary complex and render the protein ineffective by locking the complex into a non-functional state. An example in this category include the interfacial binding inhibitor BFA binds at the protein–protein interface of ARF1–GDP–Sec7 complex and inhibits conformational changes in ARF1 necessary for Sec7 to displace the GDP molecule [18]. Protein–protein interaction study between human ACE2 and SARS-CoV-2 S protein can be used to design interface binding inhibitors which can disrupt the protein–protein interaction and inhibit the attachment and entry of the virus into the host cell. In this study, we have screened a chemical compound library to screen novel drug-like molecules which can effectively bind to the interface region between RBD of SARS-CoV-2 S protein and human ACE2 which

will help in blocking the viral attachment and entry into the host cell.

Materials and methods

Retrieval of three-dimensional crystal structures of target proteins

The three dimensional X-ray crystal structure of SARS-CoV-2 spike receptor-binding domain (RBD) bound with human ACE2 (PDB ID: 6MOJ) and the crystal structure of SARS-CoV spike RBD complexed with human ACE2 (PDB ID: 2AJF) were retrieved from Protein Data Bank (PDB) (<http://www.rcsb.org/>) at a resolution of 2.45 Å and 2.90 Å respectively.

Characterization of protein–protein interface region

The protein–protein interface statistics including interface area, interface residues and residue interactions across the interface (salt bridges, disulphide bonds, hydrogen bonds and non-bonded contacts) were determined using PDBsum program [19]. The hotspot residues in protein–protein interfaces were calculated using alanine scanning mutagenesis program of DrugScorePPI webserver [20].

Sequence alignment and structural superposition

The pairwise sequence alignment between SARS-CoV-2 spike RBD and SARS-CoV spike RBD was carried out using the CLUSTAL W program [21]. The root mean square deviation (RMSD) between structures of SARS-CoV-2 spike RBD and SARS-CoV spike RBD was calculated using UCSF Chimera [22].

Retrieval of three-dimensional structures of ligands

A random set of 1,36,191 chemical compounds were retrieved from the ZINC database [23]. The chemical structures were downloaded and saved in SDF format for the virtual screening process.

Virtual screening of small drug-like molecules

The chemical library was subjected to virtual screening procedure based on selected drug-like filters such as Veber's rule [24], Lipinski's rule of five (ROF) [25] and *in silico* toxicity parameters such as mutagenicity, tumorigenicity, irritancy and reproductive effects. The physicochemical properties of the compounds were calculated using the OSIRIS DataWarrior program version 5.0 [26].

Molecular docking studies

Virtual screening of potential hits binding to SARS-CoV-2 spike RBD was performed using ArgusLab docking program version 4.0 [27]. A grid box of size 40 × 40 × 40 Å with a grid resolution of 0.4 was placed at the centre around the interface residues. The other parameters include AScore as scoring function and the Dock-Engine was set to exhaustive search. The top 100 hits were docked within the interface region using AutoDock Vina program which utilizes sophisticated gradient optimization method [28]. The target protein was prepared by removing the heteroatoms such as ions, water molecules and cocrystal ligands and the addition of polar hydrogen atoms and Kollman charges. Similarly, the chemical compounds were prepared by adding hydrogen atoms, Gasteiger charges and optimally defining the torsions. A grid box encompassing the interface region was centred at XYZ coordinates of –36.223, 27.8467 and 11.8760 with dimensions of X: 27.5694 Å, Y: 45.4190 Å and Z: 30.0170 Å. The binding conformations were clustered and

ranked according to their binding affinities. The molecular interactions (hydrogen bonds and hydrophobic interactions) between the target protein and compounds were evaluated using LigPlot + program version 1.4.5 [29].

Molecular dynamics simulation

The molecular dynamics of the spike-ligand docked complexes were performed using AMBER16 program available on Ligand and Receptor Molecular Dynamics (LARM-D) (<http://chemistry.cnu.edu.cn/ccb/server/LARM-D/>). The interactional binding mode (Int.mod) was chosen for 4 ns MD simulation in an explicit water model [30]. The steps involved in the MD simulations consist of (a) assignment of bcc charges to the ligand atoms using the antechamber module of the AMBER16 program [31] (b) generation of the coordinate and topology files of the complex using Tleap module of the AMBER16 program (c) The force fields selected for amino acid residues and ligands include the AMBER ff14SB force field [32] and General AMBER Force Field (GAFF) [33,34] respectively (d) Addition of counter ions (Na^+ and Cl^-) to the system (e) the systems were explicitly solvated in an octahedral box of TIP3P waters [35] (f) energy minimization with 2000 steps steepest descent method and 3000 steps conjugated gradient method using sander module in the AMBER16 program [36] (g) the systems were gradually heated by raising the temperature from 10 to 300 K for 30 ps. All the atoms were relaxed using periodic boundary conditions at a temperature of 300 K and pressure of 1 atm.

Analysis of trajectories

The geometric properties of the systems such as root mean square deviation (RMSD), radius of gyration (Rg) and fraction of native contacts (Q) values were analysed using the Cpptraj module in the AMBER16 program module [37]. R package was used to analyse the RMSD distribution of protein and ligand in a histogram representation. The non-native contacts were calculated using MDTraj and Bio3d was used to study the principal component analysis PCA and residue cross-correlation [38].

Binding free energy analysis

The binding free energy (ΔG_{bind}) was calculated using the following Eq. (1)

$$\Delta G_{\text{bind}} = \Delta E_{\text{bind}} - T\Delta S_{\text{sol}} - T\Delta S_{\text{conf}} \quad (1)$$

where ΔE_{bind} is the binding energy, $T\Delta S_{\text{sol}}$ is the solvation entropy and $T\Delta S_{\text{conf}}$ is the conformational entropy. While the enthalpy was calculated using the Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) or molecular mechanics generalized Born surface area (MM-GBSA) method [39], entropy was derived using an empirical method [40,41].

Results and discussion

Characterization of the interface region between ACE2 and spike protein

The human angiotensin-converting enzyme 2 (ACE2) and SARS-CoV-2 spike receptor-binding domain (RBD) and human ACE2 and SARS-CoV spike RBD were analysed for their interface regions and key residues contributing to the binding recognition between the interacting protein partners. The interface region between human ACE2 and SARS-CoV spike RBD complex was investigated and there are nineteen interface residues (Gln24, Thr27, Phe28, Lys31, His34, Glu37, Asp38, Tyr41, Gln42, Leu45, Met82, Tyr83,

Gln325, Glu329, Asn330, Lys353, Gly354, Asp355 and Arg357) of ACE2 and fifteen residues of spike RBD (Arg426, Tyr436, Tyr440, Tyr442, Leu472, Asn473, Tyr475, Asn479, Gly482, Tyr484, Thr486, Thr487, Gly488, Ile489 and Tyr491) forming an interface area of 832 Å² and 872 Å² respectively (Fig. 1A). The complex is stabilised by 1 salt bridge, 9 hydrogen bonds and 94 non-bonded contacts. Further, alanine scanning mutagenesis of the interface residues was carried out to determine residues (hotspots) contributing more to the binding between the two interacting proteins. Hot spots are residues which contributes a significant increase in the binding free energy of at least 2.0 kcal/mol when replaced with alanine [42]. Both ACE2 receptor and SARS-CoV spike RBD possess two hotspot residues in their interface regions which are Asp38 and Tyr41 in ACE2 and Tyr475 and Tyr491 in spike RBD (Suppl. Fig. S1A). The interface statistics of ACE2 and SARS-CoV-2 spike RBD complex show that there are twenty residues (Gln24, Thr27, Phe28, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Tyr41, Gln42, Leu79, Met82, Tyr83, Asn330, Asp335, Lys353, Gly354, Arg357 and Arg393) belonging to human ACE2 and seventeen residues of RBD (Lys417, Gln443, Gly446, Tyr449, Tyr453, Leu455, Phe456, Ala475, Phe486, Asn487, Tyr489, Gly496, Gln498, Thr500, Asn501, Gly502 and Tyr505) contributing to interface area of 825 Å² and 863 Å² respectively (Fig. 1B). The protein–protein complex between ACE2 and SARS-CoV-2 spike RBD were stabilized by 1 salt bridge, 9 hydrogen bonds and 101 non-bonded contacts. Three hotspot residues (Asp38, Tyr41 and Tyr83) were identified in ACE2 and two hot spot residues (Tyr489 and Tyr505) in SARS-CoV-2 spike RBD (Suppl. Fig. S1B). The interface statistics differences in terms of the number of interface residues and non-bonded contacts between the two protein–protein complexes may plausibly explain the differential nature of binding of SARS-CoV-2 and SARS-CoV spike protein to ACE2 receptor. It has been revealed from an earlier study that ACE2 binds to SARS-CoV-2 ectodomain with approximately 15 nM affinity which is around 10–20 fold higher than its binding to SARS-CoV [43].

Sequence alignment and structural superposition

The primary sequence alignment of SARS-CoV-2 and SARS-CoV spike RBD domains show a sequence similarity of 72.38% and out of the five critical residues (Tyr442, Leu472, Asn479, Thr487 and Tyr491) at the SARS-CoV spike RBD and ACE2 interface, only Tyr491 is conserved (Fig. 2A). The results corroborates with the previous studies by Xie and Chen [44]. It is worth mentioning here that Asn479 and Asn487 are two important residues in RBD which determine SARS progression and tropism, and their mutations may increase animal-to-human or human-to-human transmission [12]. Further, their structural superposition exhibits a root mean square deviation (RMSD) of 0.674 Å (Fig. 2B) which means that the structure of the RBD domain has undergone significant change during viral co-evolution within the host cell.

Virtual screening of drug-like molecules

A total of 1,36,191 chemical compounds from the ZINC database were subjected to virtual screening procedure and the molecules were filtered based on their physicochemical properties and toxicity profile. A final set of 42,889 molecules qualified the rule of five, Veber's rule and found to have favourable drug-like properties such as being non-mutagenic, non-tumorigenic, non-irritant and no adverse effects on the reproductive health.

Molecular docking studies

An initial virtual screening was performed using ArgusLab docking program considering interface region of SARS-CoV-2 spike RBD as the binding pocket for a set of 42,889 drug-like molecules. Out

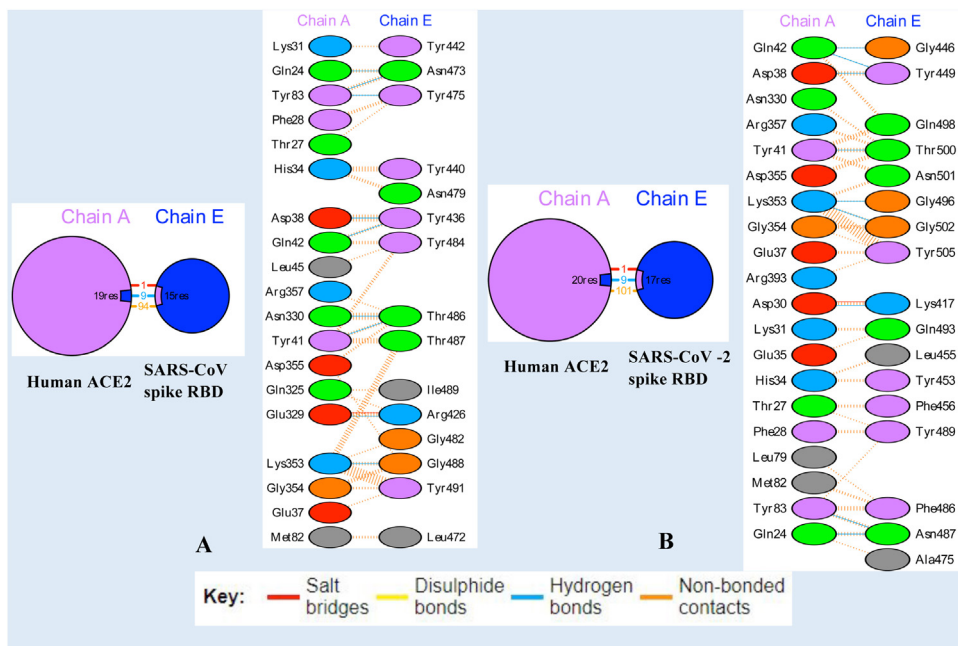


Fig. 1. The interface statistics for (A) SARS-CoV spike RBD and human ACE2 receptor complex (B) SARS-CoV-2 spike RBD and human ACE2 receptor complex depicting the interface area, residues and molecular interactions. The interface residues have been coloured according to the physicochemical properties such as positive residues (His, Lys and Arg) in blue, negative residues (Asp, Glu) in red, neutral residues (Ser, Thr, Asn, Gln) in green, aliphatic residues (Ala, Val, Leu, Ile and Met) in grey, aromatic residues (Phe, Tyr and Trp) in purple and other residues (Pro and Gly) in orange.

SARS-CoV-2	333	T N L	C P F G E V F	N A T R F A S V Y A	W N R K R I S N C V	A D Y S V	L Y N S A	S F S T F K C Y G V	382
SARS-CoV	323	.	C P F G E V F	N A T K F P S V Y A	W E R K K I S N C V	A D Y S V	L Y N S T	F F S T F K C Y G V	369
SARS-CoV-2	383	S	P T K L N D L C F	T N	V Y A D S F V I	R G D E V R Q I A P	G Q T G K I A D Y N	Y K L P D D F T G C	432
SARS-CoV	370	S	A T K L N D L C F	S N	V Y A D S F V V	K G D D V R Q I A P	G Q T G V I A D Y N	Y K L P D D F M G C	419
SARS-CoV-2	433	V I A W N S N N L D	S K	V G G N Y N Y L	Y R L	F R K S N L K	P F E R D I S T E I	Y Q A G S T P C N G	482
SARS-CoV	420	V L A W N T R N I D	A T	S T G N Y N Y K	Y R V	L R H G K L R	P F E R D I S N V P	F S P D G K P C T	468
SARS-CoV-2	483	V E G	F N C Y F P L	Q S Y G F Q P T	N G	V G V	Q P Y R V V V	L S F E L L H A P A	T V C G 526
SARS-CoV	469	P P A	L N C Y W P L	N D Y G F Y T T	T G	I G V	Q P Y R V V V	L S F E	502

A



B RMSD=0.674 Å

Fig. 2. (A) Pairwise sequence alignment between SARS-CoV-2 spike RBD and SARS-CoV spike RBD. The blue encircled residues indicate the five critical interface residues (Tyr442, Leu472, Asn479, Thr487 and Tyr491) where blue asterisk show the conserved interface residue (Tyr491) and the red arrow indicates the amino acid positions (Asn479 and Thr487) prone to mutations which determine animal-to-human or human-to-human transmission. (B) Structural superposition between SARS-CoV-2 spike RBD (cyan) and SARS-CoV spike RBD (tan).

of which a total of 100 high ranked hits were obtained which possess higher docking scores (docking scores ranging between -11.9152 kcal/mol and -13.7791 kcal/mol) to the interface region of spike RBD as compared to the control Emodin (-8.2187 kcal/mol) (Suppl. Table S1). Emodin is a naturally occurring anthraquinone

derivative extracted from *Rheum officinale* and *Polygonum multiflorum* which has been demonstrated to have anti-SARS activity by blocking the binding of Spike (S) protein to ACE2 and reducing the infectivity of S protein-pseudotyped retrovirus to Vero E6 cells [45]. The 100 high ranked hits were further taken for

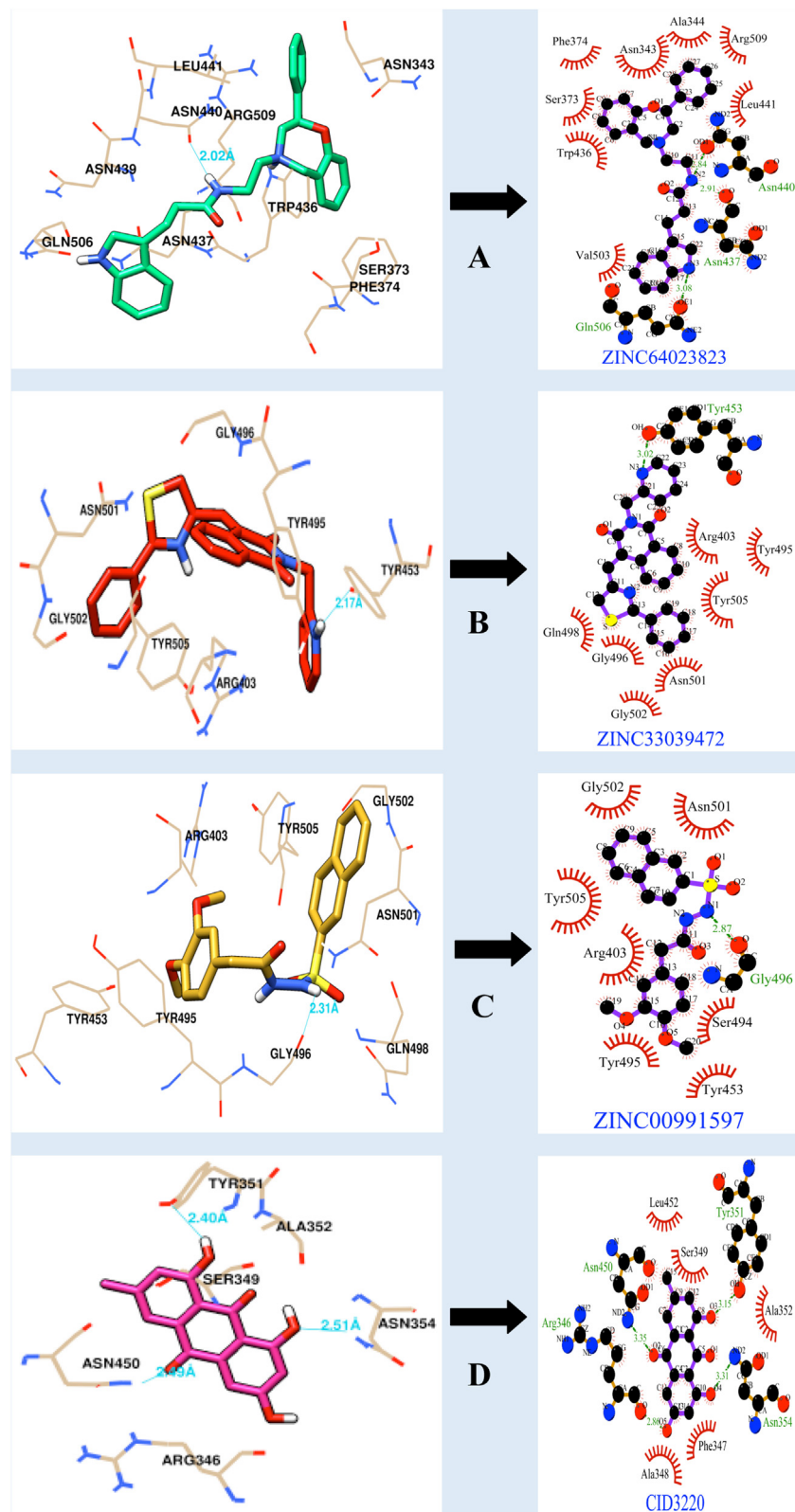


Fig. 3. Binding conformations and molecular interactions between SARS-CoV-2 spike RBD and (A) ZINC64023823 (B) ZINC33039472 (C) ZINC00991597 (D) Emodin (CID3220). The binding conformations on the left side panel show the molecules in stick representations surrounded by interacting residues in wireframe with hydrogen bonds depicted in cyan. The LigPlot+ results show the molecular interactions in two-dimensional representation with hydrogen bonds depicted in green and residues contributing to hydrophobic interactions are rendered with red arcs with protruding eyelashes.

the second round of virtual screening with Autodock vina program for rescoring and reranking the hits (Suppl. Table S2). The top 3 hits obtained were ZINC64023823 (−8.2 kcal/mol),

ZINC33039472 (−8.0 kcal/mol) and ZINC00991597 (−7.9 kcal/mol) which possessed higher binding affinity to SARS-CoV-2 spike RBD as compared to emodin (−6.1 kcal/mol) (Table 1). The drug-like

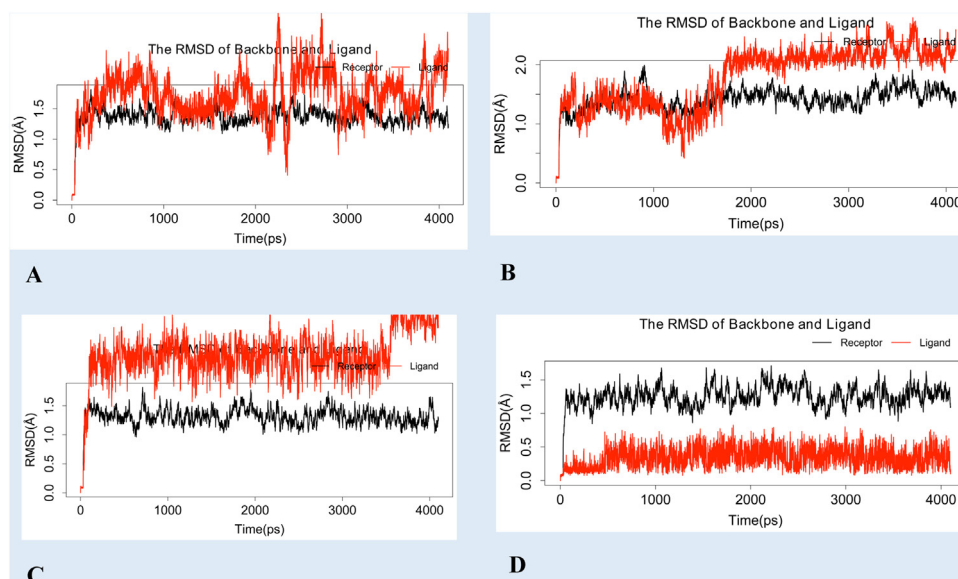
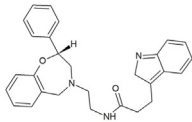
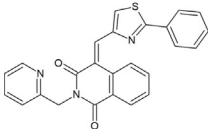
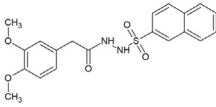
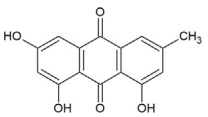


Fig. 4. Plot of root mean square deviation (RMSD) versus simulation time (ps) for (A) spike(RBD).ZINC64023823 (B) spike(RBD).ZINC33039472 (C) spike(RBD).ZINC00991597 (D) spike(RBD).Emodin. The black line indicates the RMSD curve of backbone atoms of protein whereas the red line corresponds to RMSD curve of heavy atoms of ligand in each plot.

Table 1

The binding energies of top 3 ranked molecules along with the control (Emodin) obtained through virtual screening using AutoDock Vina docking program.

Molecules	Structure	IUPAC name	Binding energy (kcal/mol)
ZINC64023823		3-((2R)-2-phenyl-3,5-dihydro-2H-1,4-benzoxazin-4-yl)ethylpropanamide	−8.2
ZINC33039472		(4E)-4-((2-phenylthiazol-4-yl)methylene)-2-(2-pyridylmethyl)isoquinoline-1,3-dione	−8.0
ZINC00991597		2-(3,4-dimethoxyphenyl)-N-(2-naphthylsulfonyl)acetohydrazide	−7.9
Emodin (CID3220)		1,3,8-trihydroxy-6-methylanthracene-9,10-dione	−6.1

properties of the top 3 compounds along with the control are enumerated in Table 2. The interaction between ZINC64023823 and spike protein is strengthened by three hydrogen bonds with residues Asn437, Asn440 and Gln506 and hydrophobic interactions with eight residues including Leu441, Arg509, Ala344, Asn343, Phe374, Ser373, Trp436 and Val503 (Fig. 3A). Whereas, the second lead molecule ZINC33039472 forms one hydrogen bond with Tyr453 within the binding cleft of spike protein and the interaction is further reinforced through hydrophobic interactions with seven residues including Arg403, Tyr495, Tyr505, Asn501,

Gly502, Gly496 and Gln498 (Fig. 3B). The third lead molecule ZINC00991597 interacts with the spike protein by establishing one hydrogen bond with Gly496 and hydrophobic interactions contributed by seven residues including Tyr453, Ser494, Asn501, Gly502, Tyr505, Arg403 and Tyr495 (Fig. 3C). Interestingly, the control emodin establishes four hydrogen bonds with residues-Arg346, Tyr351, Asn354 and Asn450 and hydrophobic interactions via five residues such as Ala348, Phe347, Ala352, Ser349 and Leu452 (Fig. 3D).

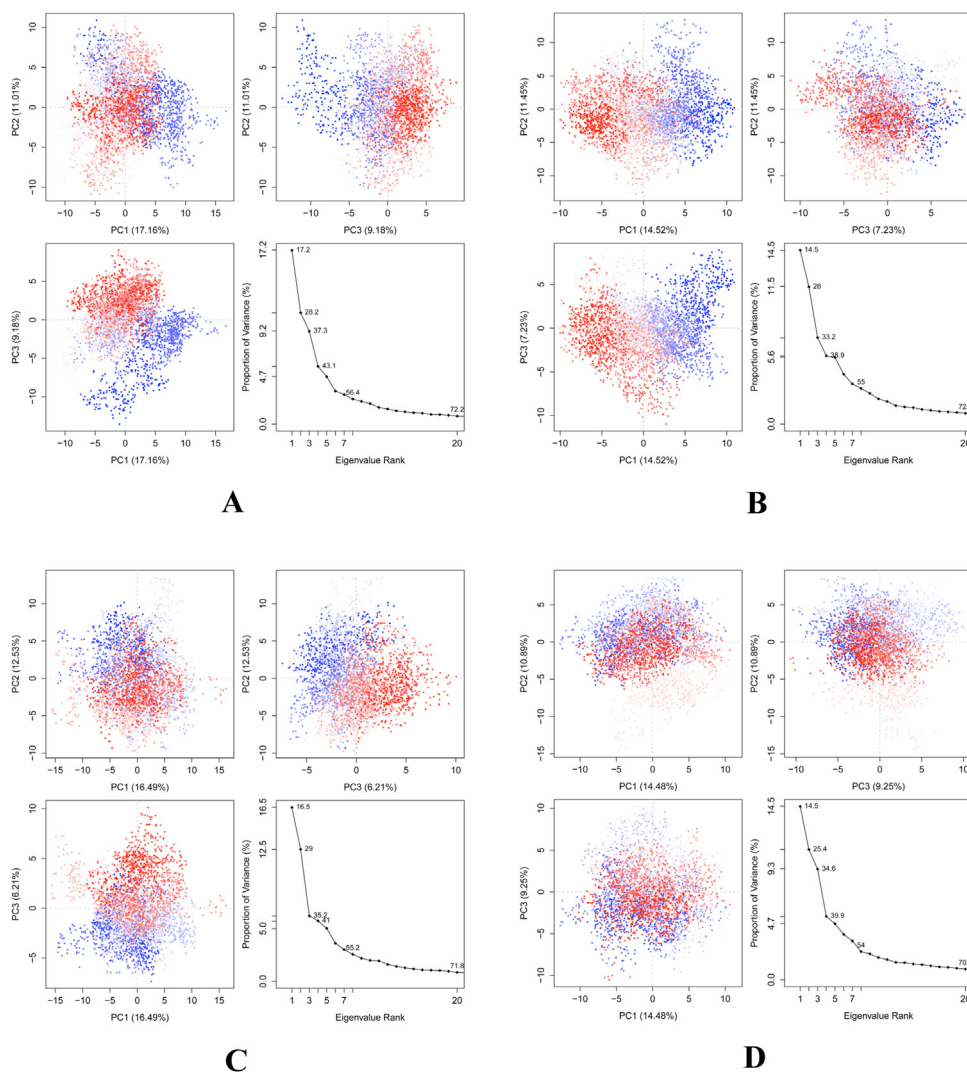


Fig. 5. Principal component analysis (PCA) for conformational changes with trajectory frames coloured from blue to white to red in order of MD simulation time for (A) spike(RBD).ZINC64023823 (B) spike(RBD).ZINC33039472 (C) spike(RBD).ZINC00991597 (D) spike(RBD).Emodin. The first twenty eigenvectors out of 582 eigenvectors (3 N, where N = 194 C α atoms) capture at least 70% of the total variance and first three eigenvectors (PC1, PC2 and PC3) contributes over 50% to the total variance.

MD simulation analysis

The binding modes of the best three lead molecules-ZINC64023823, ZINC33039472 and ZINC00991597 along with the control (emodin) were further explored using molecular dynamics simulation study in an aqueous environment for a simulation time of 4 ns. The geometric properties of the protein-ligand complexes such as root mean square deviation (RMSD), radius of gyration (Rg) and fraction of native contacts (Q) values were evaluated to check the stability of the system. The RMSD gives the measurement of root mean square deviation of atomic positions which is used to calculate the average distance between the atoms of superimposed structures of protein and ligand over a period of time [46]. The average RMSD of C α atoms of spike protein and heavy atoms of ZINC64023823 in spike(RBD).ZINC64023823 complex was found to be 1.37 ± 0.1644 Å and 1.734 ± 0.3857 Å respectively with respect to the starting structures (Fig. 4A). In the case of spike(RBD).ZINC33039472 complex, the average RMSD of C α atoms of spike protein and heavy atoms of ZINC33039472 were 1.414 ± 0.2028 Å and 1.78 ± 0.5061 Å respectively (Fig. 4B). The average RMSD of C α atoms of spike protein and heavy atoms of ZINC00991597 in spike(RBD).ZINC00991597 complex was found to be 1.308 ± 0.1676 Å and 2.374 ± 0.4118 Å (Fig. 4C).

Whereas, spike(RBD).emodin complex has an average RMSD of 1.25 ± 0.1744 Å for C α atoms of spike protein and 0.3439 ± 0.1491 Å for heavy atoms of emodin (Fig. 4D). Rg can be explained as the root mean square distance from each atom of the system to its centre of mass [47]. The Rg values for protein-ligand complexes: spike(RBD).ZINC64023823 and spike(RBD).ZINC33039472 show stable fluctuation between 18.1 to 18.5 Å whereas spike(RBD).ZINC00991597 and spike(RBD).emodin complexes stably fluctuated between 18.0 to 18.4 Å (Suppl. Fig. S2). The Q gives an idea about conformational dynamics and the transition states of a protein with a folding free energy barrier [48]. The Q values for ZINC64023823, ZINC33039472 and ZINC00991597 along with the control (emodin) were found to be 0.976819 ± 0.006382 , 0.965613 ± 0.013479 , 0.983011 ± 0.008702 and 0.957289 ± 0.012163 respectively (Suppl. Fig. S3). Principal component analysis (PCA) or essential dynamics (ED) is considered as a powerful method for clustering the conformations of a protein and extracting the large concerted modes of fluctuations from trajectories of MD simulations. The contribution of eigenvector 1 (PC1) towards the total mean square fluctuations were found to be 20.916 Å² (17.158%), 18.749 Å² (14.518%), 19.777 Å² (16.486%) and 16.677 Å² (14.476%) for spike(RBD).ZINC64023823, spike(RBD).ZINC33039472, spike(RBD).ZINC00991597 and spike(RBD).emodin respectively



Fig. 6. Per-residue binding energy analysis showing the top ten residues contributing significantly to the total binding free energy of two systems-(A) spike(RBD).ZINC33039472 (B) spike(RBD).Emodin.

(Fig. 5). Eigenvector 2 contributions to the total mean square fluctuations for spike(RBD).ZINC64023823, spike(RBD).ZINC33039472, spike(RBD).ZINC00991597 and spike(RBD).emodin were calculated to be 13.421 Å² (11.009%), 14.789 Å² (11.451%), 15.030 Å² (12.529%) and 12.549 Å² (10.891%) respectively. Whereas eigenvector 3 (PC3) also shares significant contributions to the total mean square fluctuations in spike(RBD).ZINC64023823, spike(RBD).ZINC33039472, spike(RBD).ZINC00991597 and spike(RBD).emodin complexes with their corresponding eigenvalues as 11.192 Å² (9.182%), 9.341 Å² (7.233%), 7.448 Å² (6.209%) and 10.663 Å² (9.255%).

Binding free energy analysis

The binding free energies of ZINC64023823 ($\Delta PB = -2.91$ kcal/mol, $\Delta GB = -3.00$ kcal/mol), ZINC33039472 ($\Delta PB = -9.08$ kcal/mol, $\Delta GB = -11.73$ kcal/mol), ZINC00991597 ($\Delta PB = -2.53$ kcal/mol, $\Delta GB = -3.67$ kcal/mol) and the control (emodin) ($\Delta PB = -6.02$ kcal/mol, $\Delta GB = -6.41$ kcal/mol) were determined using Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and molecular mechanics generalized Born surface area (MM-GBSA) methods. (Table 3). Out of the three lead compounds, ZINC33039472 has a binding free energy value higher than the control with major contribution by Gas phase energy (-33.89 ± 3.56 kcal/mol) and van der Waals energy (-30.20 ± 3.89 kcal/mol). The top ten residues contributing towards the bind-

ing interaction between ZINC33039472 and spike protein include Arg403 (-2.43 kcal/mol), Gln493 (-1.12 kcal/mol), Tyr505 (-1.07 kcal/mol), Phe497 (-0.69 kcal/mol), Leu455 (-0.66 kcal/mol), Gly496 (-0.61 kcal/mol), Tyr495 (-0.6 kcal/mol), Tyr453 (-0.31 kcal/mol), Ser494 (-0.27 kcal/mol) and Glu406 (-0.22 kcal/mol) (Fig. 6A). In case of spike(RBD).emodin complex, residues such as Tyr449 (-2.29 kcal/mol), Phe490 (-1.77 kcal/mol), Leu452 (-0.62 kcal/mol), Leu492 (-0.27 kcal/mol), Gln493 (-0.19 kcal/mol), Pro491 (-0.08 kcal/mol), Ile472 (-0.07 kcal/mol), Ser494 (-0.06 kcal/mol), Thr470 (-0.06 kcal/mol) and Tyr451 (-0.04 kcal/mol) contribute significantly to the total binding energy (Fig. 6B). In both the cases, the top ten residues have more contribution of gas phase energy and van der Waals energy to the total binding energy as compared to the electrostatic energy contribution.

Discussion

The high mortality rates of COVID-19 and lack of effective approved therapeutics remains a global concern. COVID-19 is caused by a positive-stranded RNA virus known as severe acute respiratory syndrome coronavirus-2. The virus makes use of the spike protein receptor-binding domain (RBD) to gain entry into the host cell by recognizing and binding to the host angiotensin-converting enzyme 2 (ACE2) receptor. Blocking the protein–protein interaction between ACE2 and spike protein receptor can be envisioned as a novel approach to prevent the virus from entering

Table 2

Drug-like physicochemical profile of top 3 lead molecules along with the control.

Molecule	MW ^a	cLogP ^b	cLogS ^c	HBA ^d	HBD ^e	TPSA ^f (Å ²)	RB ^g	Druglikeness	Mutagenic	Tumorigenic	Reproductive effective	Irritant
ZINC64023823	441.573	1.1227	−3.549	5	3	56.74	7	1.5967	None	None	None	None
ZINC33039472	423.495	3.6596	−3.855	5	0	91.4	4	5.92	None	None	None	None
ZINC00991597	400.454	1.7456	−3.112	7	2	102.11	6	3.8539	None	None	None	None
Emodin (CID3220)	270.239	2.3402	−4.186	5	3	94.83	0	−1.1275	High	High	High	High

a: Molecular weight; **b:** Partition coefficient between n-octanol and water; **c:** Aqueous solubility at 25° and pH = 7.5; **d:** Hydrogen bond acceptor; **e:** Hydrogen bond donor; **f:** Polar surface area; **g:** Rotatable bonds.

Table 3

MM-PB(GB)SA binding free energy calculations for top 3 lead molecules along with the control.

Molecules	ELE ^a (kcal/mol) (1)	VDW ^b (kcal/mol) (2)	GAS ^c (kcal/mol) (3)	PBSOL ^d (kcal/mol) (4)	PBTOT ^e (kcal/mol) (5 = 1+2 + 4)	GBSOL ^d (kcal/mol) (6)	GBTOT ^e (kcal/mol) (7 = 1+2 + 6)	TS ^f (kcal/mol) (8)	ΔPB ^g (kcal/mol) (9 = 5+8)	ΔGB ^h (kcal/mol) (10 = 7+8)
ZINC64023823	6.84 ± 4.41	−26.11 ± 3.4	−19.26 ± 5.9	3.78 ± 4.75	−15.49 ± 2.94	3.68 ± 4.09	−15.58 ± 3.29	12.58 ± 1.75	−2.91	−3.00
ZINC33039472	−3.69 ± 1.78	−30.20 ± 3.89	−33.89 ± 3.56	12.34 ± 1.94	−21.55 ± 3.39	9.69 ± 1.18	−24.20 ± 3.05	12.47 ± 1.78	−9.08	−11.73
ZINC00991597	−8.89 ± 3.91	−24.84 ± 2.71	−33.73 ± 5.51	18.04 ± 3.86	−15.69 ± 4.42	16.90 ± 3.2	−16.83 ± 3.43	13.16 ± 2.42	−2.53	−3.67
Emodin (CID3220)	−1.80 ± 2.09	−15.06 ± 3.61	−16.86 ± 4.77	5.35 ± 3.18	−11.52 ± 2.98	4.95 ± 2.07	−11.91 ± 3.32	5.50 ± 3.00	−6.02	−6.41

a: Electrostatic energy; **b:** van der Waals energy; **c:** Gas phase energy; **d:** Non-polar and polar solvation energy; **e:** Summation of electrostatic energy, van der Waals energy and non-polar and polar solvation energy; **f:** Entropic energy; **g, h:** Final estimated binding free energy.

the host cells and thus effectively control the pathogenesis of the disease. Hence, herein we have characterized the protein–protein interface of the SARS-CoV-2 RBD and ACE2 protein and used the information to target the spike protein interface region. The SARS-CoV-2 spike protein binding to ACE2 occurs through an interface area of 825 Å² consisting of twenty residues of RBD with high energetic contributions (hot spot residues) by Tyr489 and Tyr505. SARS-CoV-2 RBD protein sequence shows a similarity of 72.38% with that of SARS-CoV and out of five critical residues (Tyr442, Leu472, Asn479, Thr487 and Tyr491) at the spike receptor-ACE2 complex, we observed the conservation of the only Tyr491. Further, the three-dimensional structural difference between SARS-CoV-2 spike and SARS-CoV spike is subtle with an RMSD of 0.674 Å. This subtle differences in the sequence, structure and key binding residues perhaps explain the higher binding affinity of SARS-CoV-2 S protein to ACE2 as compared to SARS-CoV S protein. A freely accessible database such as ZINC database offers a large repertoire of chemical entities which have been used in the present virtual screening procedure. Out of a random set of 1,36,191 compounds, a total of 42, 889 molecules cleared the ROF, Verber's rule and *in silico* toxicity filters and were found to be orally bioactive. The filtered drug-like molecules were docked into the binding pocket within the interface region of RBD and the docking scores were compared with the control (emodin). Emodin is a natural anthraquinone derivative which is previously reported to block the SARS-CoV Spike (S) protein binding to Vero E6 cells and exhibit percent inhibition of 94.12 ± 5.90 % at 50 µM concentration against S protein–pseudotyped retrovirus infection [45]. Finally, three hits were shortlisted based on the docking scores and molecular interaction profile. The protein–ligand complexes of the three hits were further subjected to molecular dynamics simulation studies in an aqueous environment and various geometric and thermodynamic properties of the docked complexes were evaluated. MD simulation studies confirmed the stable trajectories of the protein–ligand complexes as inferred from RMSD, Rg and Q graphs. The principal component analysis highlights significant contributions of the first three eigenvectors (PCA 1–3) to the total mean square fluctuations. Further, the binding free energy calculation using MM-PBSA and MM-GBSA methods reveal that all the three hits exhibited favourable negative free energy values and the hit molecule ZINC33039472 showed higher binding affinity to spike receptor as compared to the control. Further, the energy decomposition analysis of spike(RBD).ZINC33039472 complex reveals a major contribution of gas-phase energy and van der Waal energy and the top ten residues which contributed majorly to the binding free energy include Arg403, Gln493, Tyr505, Phe497, Leu455, Tyr495, Tyr453, Ser494 and Glu406. The present work has a few limitations such as short duration of molecular dynamics simulation. Hence, it would be quite interesting to explore the dynamics of the protein–ligand complexes at a much longer time scales which would further provide better insights into the binding modes of the lead molecules at the atomic level.

Conclusion

Using protein–protein interface targeted drug design approach, we have successfully identified ZINC33039472 as the best lead molecule which shows stronger interaction with the spike receptor-binding domain (RBD) as compared to emodin and thus the molecule could be a promising lead which can effectively block the interaction between SARS-CoV-2 spike RBD and human ACE2 receptor thereby hindering the pathogenesis of COVID-19. Further, the identified lead could be taken for various preclinical studies for developing into anti-coronavirus drug candidate molecules.

Funding

No funding sources.

Competing interests

None declared.

Ethical approval

Not required.

Acknowledgements

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of the research through the research group project #RG-1438-015.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jiph.2020.12.014>.

References

- [1] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581:215–20.
- [2] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497–506.
- [3] Wang N, Shang J, Jiang S, Du L. Subunit vaccines against emerging pathogenic human coronaviruses. *Front Microbiol* 2020;11:298.
- [4] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74.
- [5] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020:1–4.
- [6] Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 2020;17:613–20.
- [7] Tang D, Comish P, Kang R. The hallmarks of COVID-19 disease. *PLoS Pathog* 2020;16:e1008536.
- [8] Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, et al. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* 2004;363:938–47.
- [9] Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;426:450–4.
- [10] Raj VS, Mou H, Smits SL, Dekkers DHW, Müller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013;495:251–4.
- [11] Du L, Kao RY, Zhou Y, He Y, Zhao G, Wong C, et al. Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity. *Biochem Biophys Res Commun* 2007;359:174–9.
- [12] Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* (80-) 2005;309:1864–8.
- [13] Ryan DP, Matthews JM. Protein–protein interactions in human disease. *Curr Opin Struct Biol* 2005;15:441–6.
- [14] Gonzalez MW, Kann MG. Protein interactions and disease. *PLoS Comput Biol* 2012;8:e1002819.
- [15] Gurung AB, Bhattacharjee A, Ali MA, Al-Hemaid F, Lee J. Binding of small molecules at interface of protein–protein complex—A newer approach to rational drug design. *Saudi J Biol Sci* 2017;24:379–88.
- [16] Shi A, Murai MJ, He S, Lund G, Hartley T, Purohit T, et al. Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood J Am Soc Hematol* 2012;120:4461–9.
- [17] Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, et al. RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* 2010;464:431–5.

- [18] Mossesso E, Corpina RA, Goldberg J. Crystal structure of ARF1•Sec7 complexed with Brefeldin A and its implications for the guanine nucleotide exchange mechanism. *Mol Cell* 2003;12:1403–11.
- [19] Laskowski RA, Jabłońska J, Právda L, Vařeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. *Protein Sci* 2018;27:129–34.
- [20] Krüger DM, Gohlke H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein–protein interactions. *Nucleic Acids Res* 2010;38:W480–6.
- [21] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–80.
- [22] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12, <http://dx.doi.org/10.1002/jcc.20084>.
- [23] Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–82, <http://dx.doi.org/10.1021/ci049714+>.
- [24] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 2002;45:2615–23.
- [25] Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* 2004;1:337–41, <http://dx.doi.org/10.1016/j.ddtec.2004.11.007>.
- [26] Sander T, Freyss J, von Korff M, Rufener C. DataWarrior: an open-source program for chemistry aware data visualization and analysis. *J Chem Inf Model* 2015;55:460–73, <http://dx.doi.org/10.1021/ci500588j>.
- [27] Thompson MA. Molecular docking using ArgusLab, an efficient shape-based search algorithm and the AScore scoring function Philadelphia. *ACS Meet* 2004;vol. 172. p. 42.
- [28] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31:455–61, <http://dx.doi.org/10.1002/jcc.21334>.
- [29] Laskowski RA, Swindells MB. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J Chem Inf Model* 2011;51:2778–86, <http://dx.doi.org/10.1021/ci200227u>.
- [30] Yang J-F, Wang F, Chen Y-Z, Hao G-F, Yang G-F. LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. *Brief Bioinform* 2020;21:2206–18.
- [31] Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 2002;23:1623–41.
- [32] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;11:3696–713.
- [33] Wang B, Merz KM. A fast QM/MM (quantum mechanical/molecular mechanical) approach to calculate nuclear magnetic resonance chemical shifts for macromolecules. *J Chem Theory Comput* 2006;2:209–15.
- [34] Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem* 2004;25:1157–74.
- [35] Price DJ, Brooks III CL. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* 2004;121:10096–103.
- [36] Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–88, <http://dx.doi.org/10.1002/jcc.20290>.
- [37] Roe DR, Cheatham III TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 2013;9:3084–95.
- [38] Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Cavas LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 2006;22:2695–6.
- [39] Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 2011;51:69–82.
- [40] Hao G-F, Zhu X-L, Ji F-Q, Zhang L, Yang G-F, Zhan C-G. Understanding the mechanism of drug resistance due to a codon deletion in protoporphyrinogen oxidase through computational modeling. *J Phys Chem B* 2009;113:4865–75.
- [41] Pan Y, Gao D, Zhan C-G. Modeling the catalysis of anti-cocaine catalytic antibody: competing reaction pathways and free energy barriers. *J Am Chem Soc* 2008;130:5140–9.
- [42] Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* 2001;17:284–5.
- [43] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* (80-) 2020;367:1260–3.
- [44] Xie M, Chen Q. Insight into 2019 novel coronavirus—an updated interim review and lessons from SARS-CoV and MERS-CoV. *Int J Infect Dis* 2020;94:119–24.
- [45] Ho T-Y, Wu S-L, Chen J-C, Li C-C, Hsiang C-Y. Emodin blocks the SARS coronavirus spike protein and angiotensin-converting enzyme 2 interaction. *Antiviral Res* 2007;74:92–101.
- [46] Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol* 1994;235:625–34.
- [47] Lobanov MY, Bogatyreva NS, Galzitskaya OV. Radius of gyration as an indicator of protein structure compactness. *Mol Biol* 2008;42:623–8.
- [48] Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci* 2013;110:17874–9.